Gobierno de la Provincia de Buenos Aires

Local Crime Modeling with Geographic Weighted Regression (Space Varying Relationships)

Police Confrontations Example

Gastón Pezzuchi, MSc
Luis Castro, PhD

Gobierno de la Provincia de Buenos Aires

Police Confrontations & Violent Offenders

1999 - 2001

Gobierno de la Provincia de Buenos Aires

If a lunatic scribbles a jumble of mathematical symbols it does not follow that the writing means anything merely because to the inexpert eye it is indistinguishable from higher mathematics.

Eric Temple Bell (Newman 1856, 308).

If a naïve researcher completes a standard statistical analysis of georeferenced data, it does not follow that the data analytic results have turned data into meaningful information merely because to the inexpert eye they are indistinguishable from conventional statistics results!

Daniel A. Griffith, Larry J. Lane (March 1998).

Gobierno de la Provincia de Buenos Aires

Argentinean Republic
Area: 3,761,274 Km²
Population: 36,223,947

Buenos Aires Province
Area: 307,571 Km2
Population: 13,818,677

Gobierno de la Provincia de Buenos Aires

Apologies

- For the Grammar & Spelling mistakes
  - (our native tongue is SPANISH)

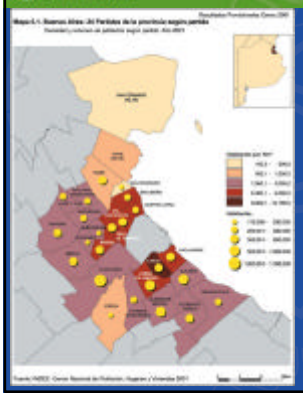
Gobierno de la Provincia de Buenos Aires

Study Area

Basic Data

	Population	Area [km²]	Density [Hab / km²]	Police Beats[4]	Police Officers	Police Officers by Beats / 1000 inhab.
República Argentina[2]	36,223,947	2,791,810	13	NO DATA	NO DATA	NO DATA
Ciudad Autónoma de Buenos Aires[3]	2,768,772	200	13,844	53	10,200	3.7
Provincia de Buenos Aires	13,818,677	307,571	45	335	19,107	1.4
Conurbano Bonaerense[4]	8,684,953	3,630	2,393	168	9,780	1.1
Resto de la Provincia[5]	5,133,724	303,941	17	167	9,327	1.8

[1] Only Police Beats, not decentralized Units.
[2] The are of the "República Argentina" (emerged lands) is 3.761.274 km² , 2.791.810 km² of them are on South America, y 9.464 km² are on the Antartica, including the "Islas Orcas del Sur", "Georgias del Sur" and "Sándwich del Sur).-
[3] Capital de la República Argentina.-
[4] INDEC's definition,it includes only **24 counties**, and not the "expanded" definition with **29 partidos.**
[5] The remaining 110 counties (there are 134 counties in Buenos Aires Province).



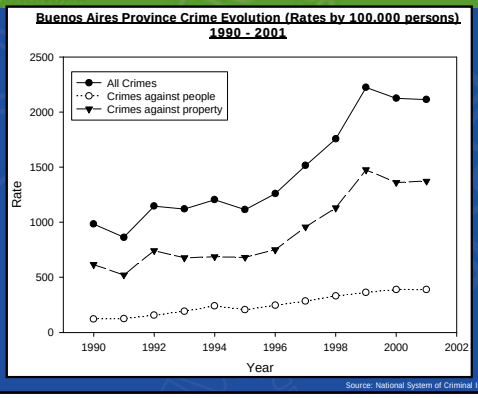
Density and Volume of Population
By County – 2001 Census

What is a Police Confrontation?

- Essentially a shooting,
- Police Officer involved,
- As a result of Police Activity.
- It is registered even if the Police Officer doesn't have the opportunity to defend himself.-

Context

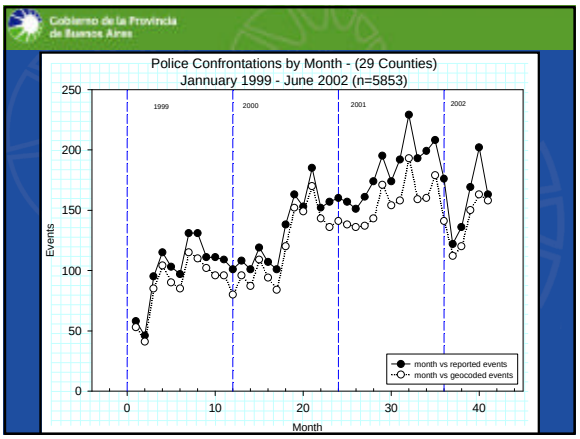
- Increase in crime rates over the last decade, (+100 %)
- Increase in urban violence,
- Follow up on previous studies.
  - Spatial Point Pattern Analysis – ESDA Techniques
  - See for example:
    - *Spatial Statistics Analysis of Police Confrontations in a Large Argentinean Urban Area: Preliminary Findings - Pezzuchi G., 2002.*
    - *Spatial Analysis of Police Confrontations in a Large Argentinean Urban Setting, Pezzuchi, G; Jorge Ortiz, 2002.*

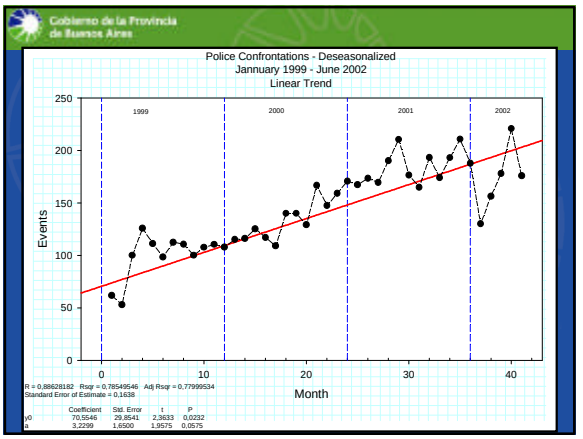


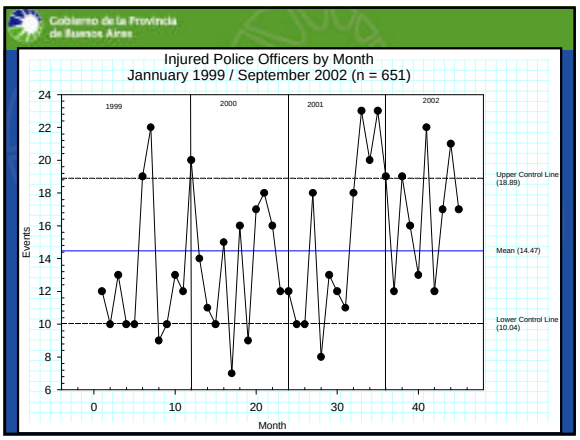
Dataset: Confrontations – 29 Counties

Dataset	Reported Events	GIS Dataset	
		Geocoded Events	Geocode Rate
1999	1208	1057	86 %
2000	1644	1481	90 %
2001	2200	1869	85 %
2000 – 2001	3844	3350	87 %
1999 – 2001	5052	4407	87 %

Point Data – (Aggregated to Areal Units (census tracts for this study))







Datasets: Offenders last known address – 29 Counties

Dataset	Recorded Addresses [1]	GIS Datasets	
		Geocoded	Geocoe Rate [2]
2000	1294	926	72 %
2001	1468	1017	70 %
2000 – 2001	2762	1943	70 %

[1] Aprox. el 10 % of each dataset included missing or incorrect data.
[2] This rate is calculated using all the recorded adresses (inlcuded the missing or incorrect data), if we had used only those records with information the geocode rate would be a much more acceptable 80-85 %.

Point Data – (Aggregated to Areal Units for this study)

Preliminary Conclusions from Previous Studies

a) There is Evidence of Clustering of events near or within settlement areas, but there is also evidence of clustering at other sites (commercial areas, major highways, etc.):

This is consistent with the genesis of this type of events, since they tend to generate along the lines of three major situations:

1) Those occurring AT the place (or quite near) of a specific crime.
2) Those occurring WHILE the police are pursuing the offenders that committed a crime.
3) Those occurring WHEN the offenders are entering into their home-area or some sort of safe zone (sanctuary).

b) The Comparison of Confrontations Events and Offender's residence shows that few settlements are located within a high volume zone.

c) More research should be done regarding the Police Deployment Strategy and its relationships with population density, overall crime rate and police risk among other factors.

More Preliminary Conclusions:

a) The Evidence of Settlements as a generators of violent offenders is at most quite weak.

b) More research is needed to study the relationship between socio-economic indicators, the distribution of violent offenders and the deployment strategies of the police force.

c) The Journey to Crime Analysis showed different regional patterns of behavior that needs to be further developed.

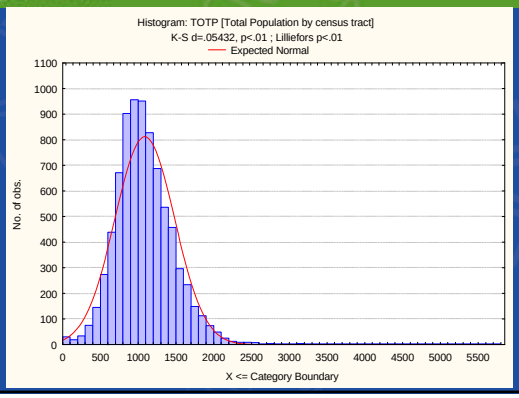
Areal Units = Census Tracts

Variable Descriptions – (24 Counties)

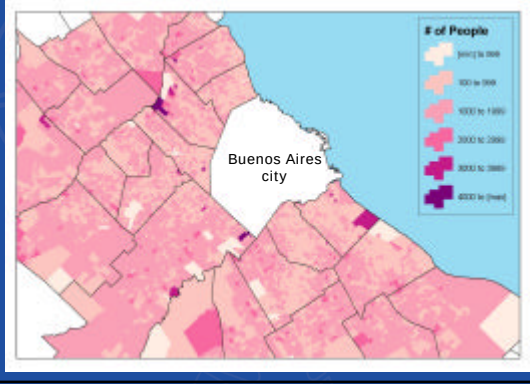
	Valid N	Mean	Minimum	Maximum	Std.Dev.
TOTP	8027	1087.305	0.000000	5788.000	394.4820
OFF_HOME	8027	0.285	0.000000	23.000	0.8017
CONF	8027	0.528	0.000000	16.000	0.9418
TOT_POL	8027	1.137	0.010000	58.829	1.9593
PROP_VILL	8027	3.166	0.000000	100.000	15.2008
WELFARE	8027	-0.018	-0.890598	1.245	0.4168
SUP_RAD	8027	0.326	0.003000	42.052	1.4642

Rounded Values where necessary

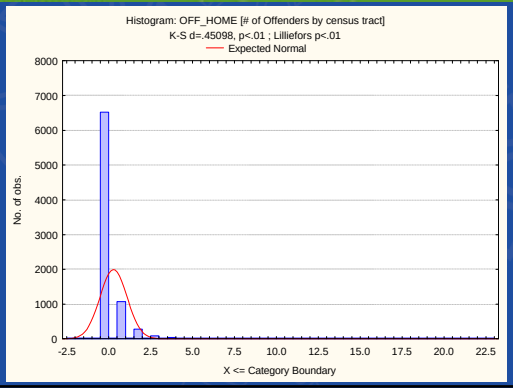
TOTP



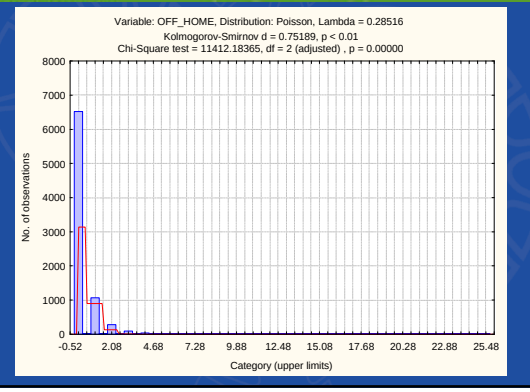
TOTP



OFF_HOME

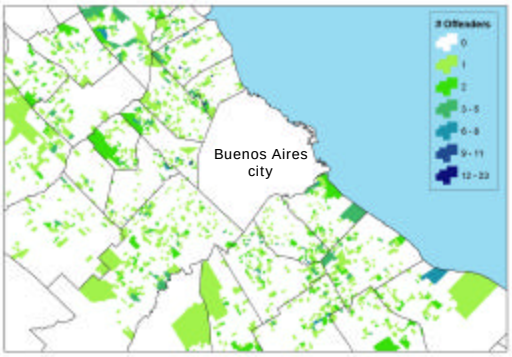


OFF_HOME



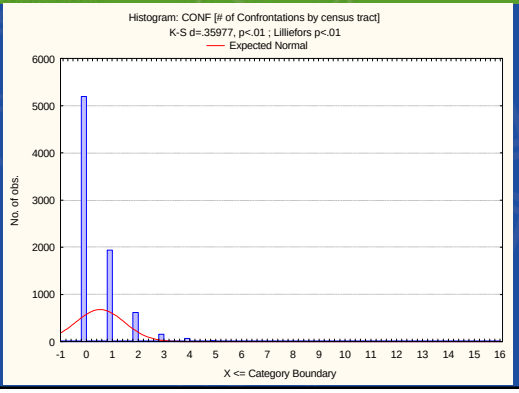
Gobierno de la Provincia de Buenos Aires

OFF_HOME



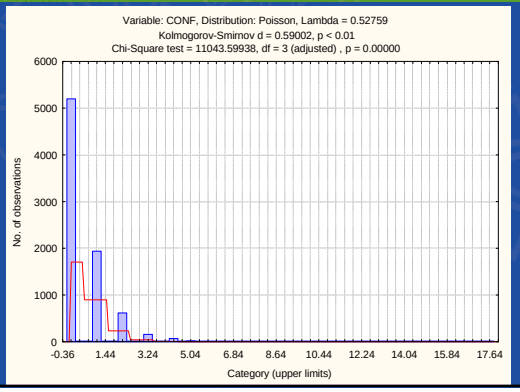
Gobierno de la Provincia de Buenos Aires

CONF



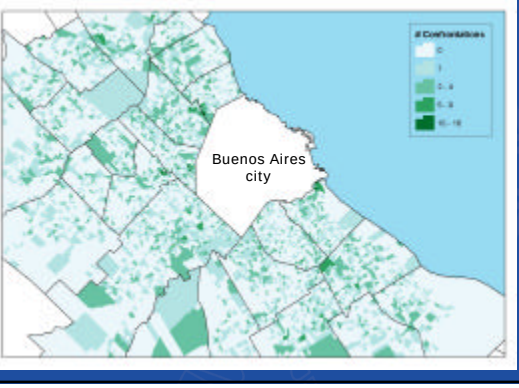
Gobierno de la Provincia de Buenos Aires

CONF



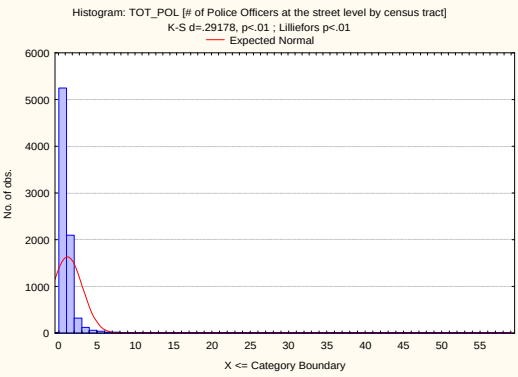
Gobierno de la Provincia de Buenos Aires

CONF



Gobierno de la Provincia de Buenos Aires

TOT_POL



Gobierno de la Provincia de Buenos Aires

Welfare Indicator

Castro / Fernandez Conti 2002 – Buenos Aires Province Ministry of Education

- How to construct a welfare indicator using 18 binary variables, and 600,000 cases:
  - 0 -> absence of the economic good
  - 1 -> possession of the good

- Using Principal Component Analysis (PCA) to **order** the observations.

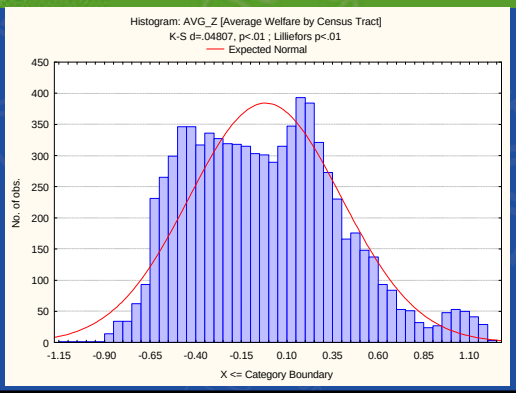
- Allows us to construct new variables as linear combinations of the original ones, maximizing the explained variance in successive axis.

- The analysis produce new uncorrelated variables.

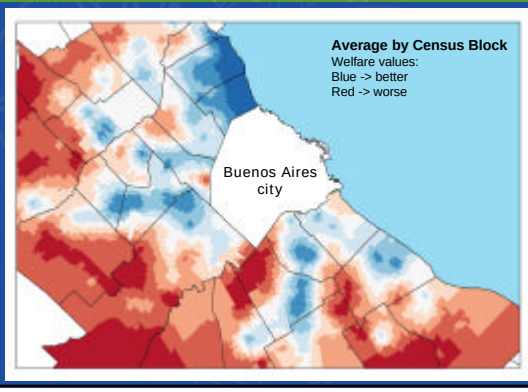
Welfare Indicator

- Our first axis orders the pupils by the number of belongings, taking into account the covariance among variables (Variables with higher variance will have higher weight in the analysis). This first axis is called the WELFARE indicator.
- The rare belongings or the too common ones will have the least influence, the median ones will provide the higher discriminator power among observations.
- The WELFARE indicator was averaged by school for this study.

WELFARE



Welfare Indicator

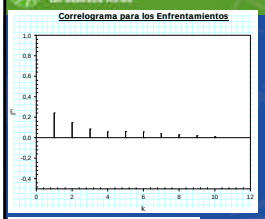


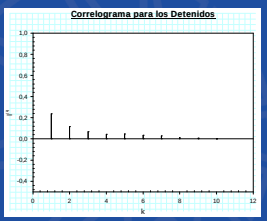
Autocorrelation table

Variable	Moran's I
[TOTP]	0.3549
[OFF_HOME]	0.1599
[CONF]	0.1539
[TOT_POL]	0.2071
[PROP_VILL]	0.5587
[WELFARE]	0.9841
[SUP_RAD]	0.3326

p: 0.001 – Queen contiguity criterion

Correlograms





Persistence of positive spatial autocorrelation -> non stationary / heterogeneity

What is GWR?

- Assume a global regression model:

$$y_i = a_0 + \sum_k a_k x_{ik} + e_i$$

- The estimator for this model is:

$$\mathbf{a} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

a := vector of global parameters to be estimated,
X := matrix of independent variables
y := vector of observations

What is GWR?

- GWR extend this framework allowing for **local parameters to be estimated**:

$$y_i = a_0(u_i, v_i) + \sum_k a_k(u_i, v_i) x_{ik} + e_i$$

(u_i, v_i) := coordinates of the ith point in space.
$a_k(u_i, v_i)$:= realization of the continuous function $a_k(u,v)$ at point *i*.

Hence we have a continuous surface of parameter values

- The GWR estimator is:

$$\mathbf{a}(u_i, v_i) = \left(\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y}$$

Where $\mathbf{W}(u_i, v_i)$ is an *n* by *n* matrix whose off-diagonal elements are zero and whose diagonal elements denote the geographical weighting of observed data *for point i*.

What is GWR? - cont

- GWR recognizes that spatial variations in relationships might exist.
- For calibration of the GWR model, it is assumed that observed data near to point *i* have more of an influence in the estimation of the $a_k(u_i, v_i)$ than do data located farther from *i*. Weighted least squares is used.
- In GWR an observation is weighted in accordance with its proximity to point *i* so that the weighting of an observation is no longer constant, but varies with *i*. The closer an observation is to *i*, more weight receives.

GWR references

- Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., (2002) - **Geographically Weighted Regression: The Analysis of Spatially Varying Relationships**, Chichester: Wiley.
- Fotheringham, A.S., Brunsdon, C., and Charlton, M.E., (2000) - **Quantitative Geography**, London: Sage

http://www.ncl.ac.uk/~ngeog/GWR/

GWR software

- R code:
  - Chris Brunsdon, Chris.Brunsdon@ncl.ac.uk
  - Luis Castro, elchinocastro@yahoo.com.ar
- MatLab code:
  - James P LeSage, (spatial econometric toolbox), www.spatial-econometrics.com
- Standalone software:
  - GWR 3.x (Fotheringham, Brunsdon, Charlton), http://www.ncl.ac.uk/~ngeog/GWR/

Violent Offenders / Confrontations Example

Why not Poisson GWR?

- Note that:
  - [OFF_HOME]:
    - Mean = 0.285163
    - Variance = 0.642694
    - Variance / Mean = 2.254
    - # of areal units with 0 counts = 6518 (81.2 % of all units)
  - [CONF]
    - Mean = 0.527594
    - Variance = 0.886947
    - Variance / Mean = 1.681
    - # of areal units with 0 counts = 5197 (64.74 % of all units)

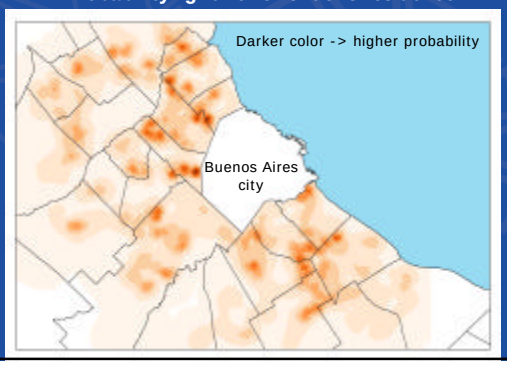
Why not GWR Poisson?

- Also:
  - GWR R Code Crashed due to memory allocation problems.
  - GWR 3.x software (Brunsdon / Fotheringham & Charlton) also crashed due to the large number of zeros. Not even extremely large bandwidths could cope with this problem.

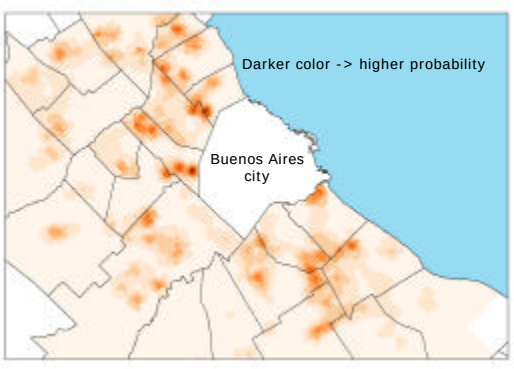
Work-around

- Construction of a "Probability" grid:
  - Using kernel density interpolation,
  - Offender's last known residence as data points,
  - We get the proportion of all incidents that occur in the grid cell. The sum of all grid cells equals a "probability" of 1.
- Aggregation of this grid to the census tract polygons -> [PROB] variable.
- Calculation of the "Probability" per unit area (due to the size variations of the census tracts) -> [PROBN] variable.

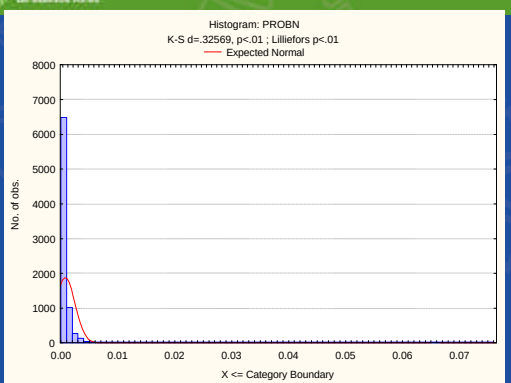
"Probability" grid for offender's residence



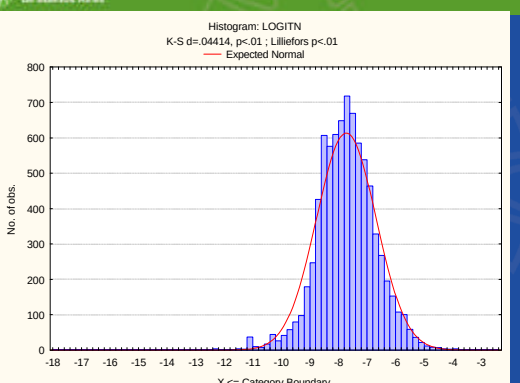
PROBN



PROBN



LOGITN



Global Model

[LOGITN] ~ [TOTP] + [CONF] + [TOT_POL] + [PROP_VI] + [WELFARE] + [SUP_RAD]

- Akaike Information Criterion (AIC): 21814.7
- R^2 (adjusted): 0.186042
- F-statistic: 306.947

Parameter	Coefficient	Std Error	t-Statistic	Probability
Intercept	-8.17164126	0.03369468	-242.520	0.0000000
Totp	0.00030646	0.00002948	10.395	0.0000000
Conf	0.22430451	0.01129345	19.861	0.0000000
Tot_pol	0.02386912	0.00727466	3.281	0.0010383
Prop_vi	0.01207260	0.00069930	17.264	0.0000000
Welfare	-0.04745940	0.02811589	-1.688	0.0914636
Sup_rad	-0.22223985	0.00980691	-22.662	0.0000000

TEST ON NORMALITY OF ERRORS

TEST	DF	VALUE	PROB
Jarque-Bera	2	2916.676	0.0000000

DIAGNOSTICS FOR HETEROSKEDASTICITY RANDOM COEFFICIENTS

TEST	DF	VALUE	PROB
Breusch-Pagan test	6	658.2302	0.0000000
Koenker-Bassett test	6	270.1695	0.0000000

SPECIFICATION ROBUST TEST

TEST	DF	VALUE	PROB
White	27	807.9657	0.0000000

GWR Model

- Local sample size: 413 (nearest neighbors) determined using AIC.
- Model fitted at data point locations.
- Results:
  - Effective number of parameters: 300.26
  - Akaike Information Criterion: 12850.28
  - R^2 (adjusted): 0.743695
  - **ANOVA TEST** (H0 := GWR model represents No improvement over a global model)

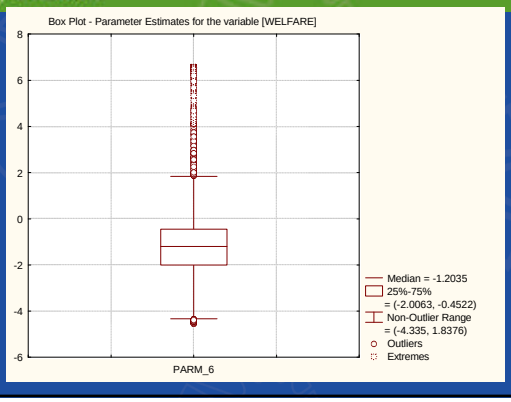
Source	SS	DF	MS	F
OLS Residuals	7103.7	7.00		
GWR Improvement	4948.6	293.26	16.8748	
GWR Residuals	2155.1	7726.74	0.2789	60.5028

GWR Model – cont.

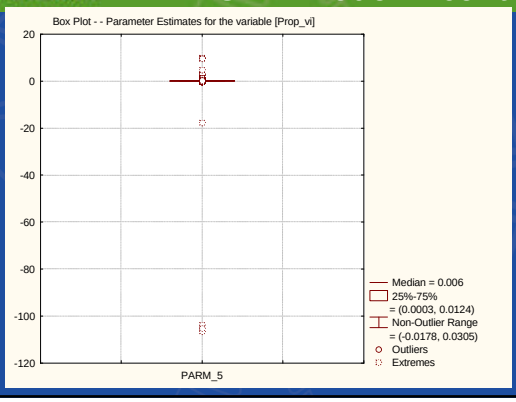
5 Number Summary of local parameter estimates

	Minimum	Lower Quartile	Median	Upper Quartile	Maximum
Intercept	-10.158063	-8.092229	-7.674777	-7.030381	-5.234107
[Totp]	-0.000837	0.000008	0.000149	0.000320	0.001068
[Conf]	-0.107722	0.027097	0.064984	0.103206	0.521091
[Tot_pol]	-1.278027	-0.075247	0.057091	0.205140	1.092299
[Prop_vi]	-106.487968	0.000260	0.006007	0.012375	9.837226
[Welfare]	-4.546778	-2.008148	-1.203934	-0.452643	6.593835
[Sup_rad]	-8.855449	-1.338788	-0.428919	-0.043107	11.461987

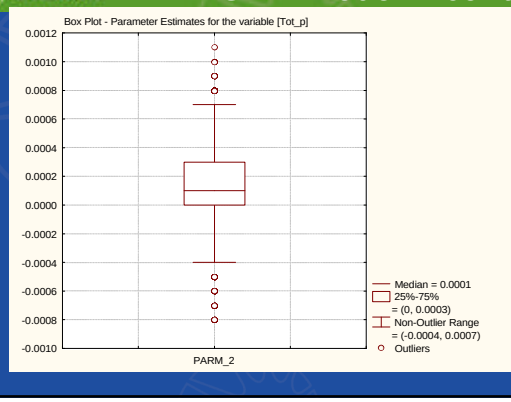
GWR Model – cont.



GWR Model – cont.



GWR Model – cont.



GWR Model – cont.

	Global Estimate		GWR Estimates	
	-1SD	+1SD	Lower Quartile	Upper Quartile
Intercept	-8.205336	-8.137947	-8.092229	-7.030381
Totp	0.000277	0.000336	0.000008	0.00032
Conf	0.213011	0.235598	0.027097	0.103206
Tot_pol	0.016594	0.031144	-0.075247	0.20514
Prop_vi	0.011373	0.012772	0.00026	0.012375
Welfare	-0.075575	-0.019344	-2.008148	-0.452643
Sup_Rad	-0.232047	-0.212433	-1.338788	-0.043107

GWR Model – cont.

Spatial Variability of Parameters

Parameter	P-value	
Intercept	0.00000	***
[Totp]	0.00000	***
[Conf]	0.49000	n/s
[Tot_pol]	0.00000	***
[Prop_vi]	0.00000	***
[Welfare]	0.00000	***
[Sup_rad]	0.00000	***

(Tests based on the Monte Carlo significance test)

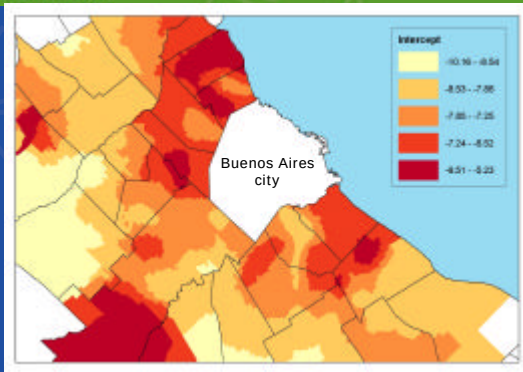
*** = significant at .1% level

GWR Model – cont.

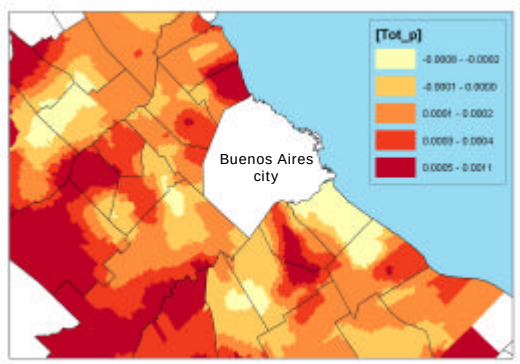
Model	Moran's I of the residuals p: 0.001
Global Model	0.7943
GWR Model	0.6391

GWR Parameter Surfaces

Intercept

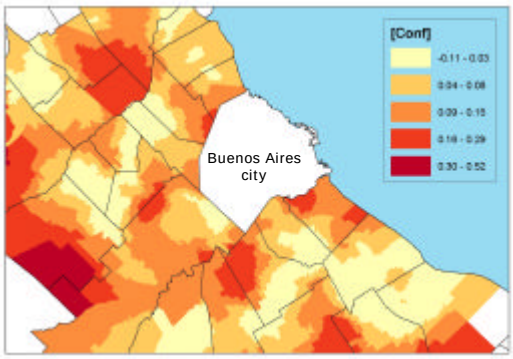


TOT_P



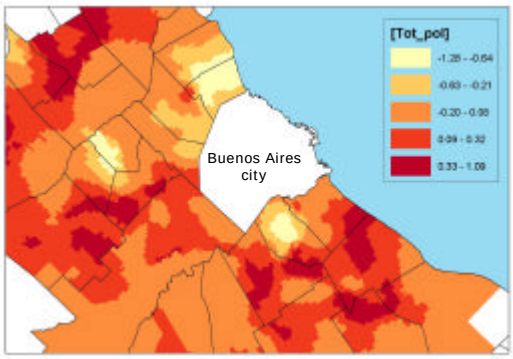
Gobierno de la Provincia de Buenos Aires

CONF



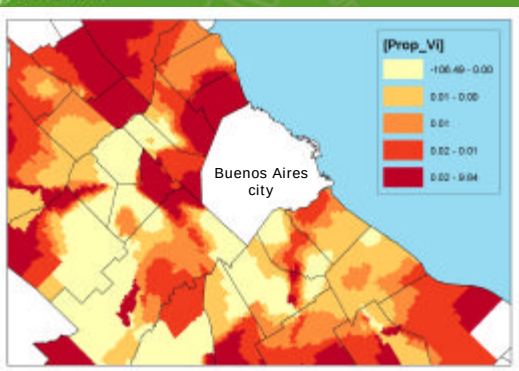
Gobierno de la Provincia de Buenos Aires

TOT_POL



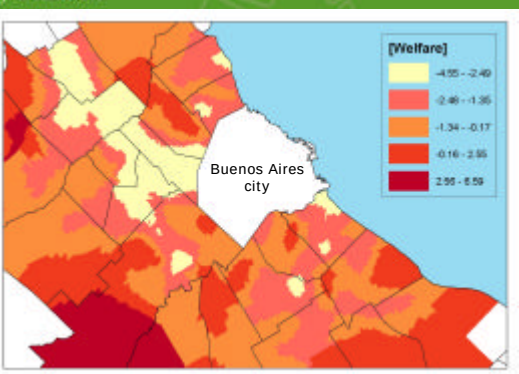
Gobierno de la Provincia de Buenos Aires

PROB_VI



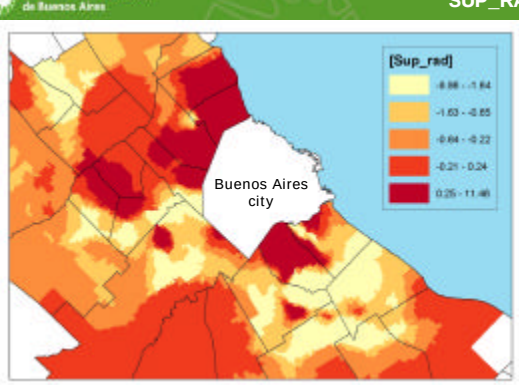
Gobierno de la Provincia de Buenos Aires

WELFARE



Gobierno de la Provincia de Buenos Aires

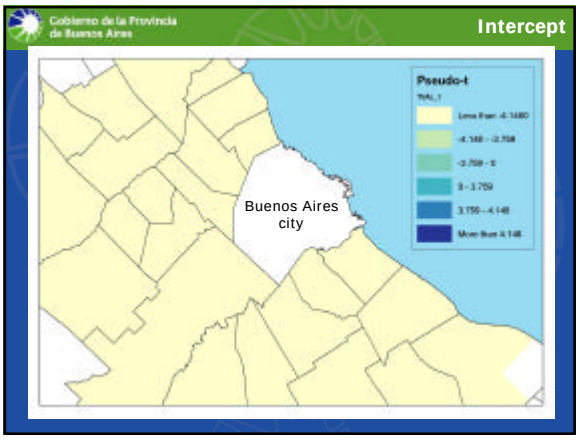
SUP_RAD

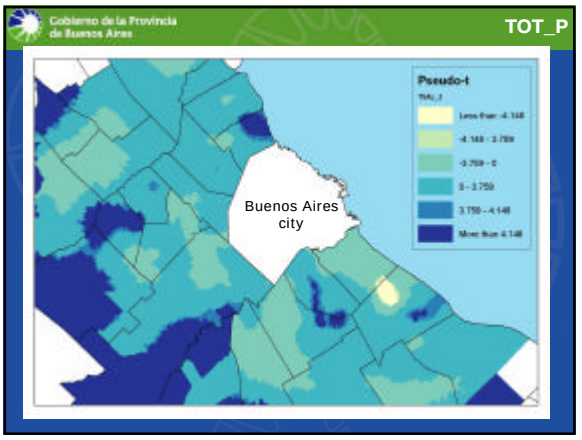


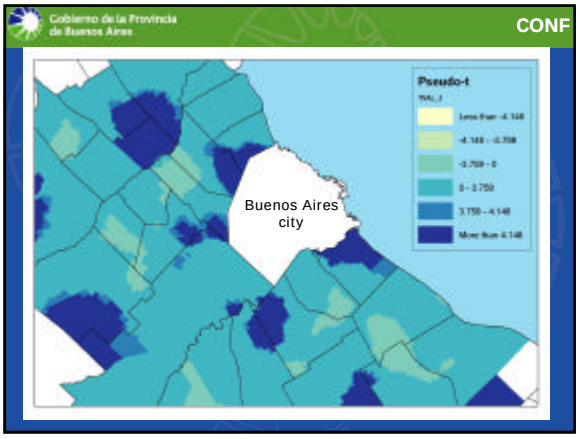
Gobierno de la Provincia de Buenos Aires

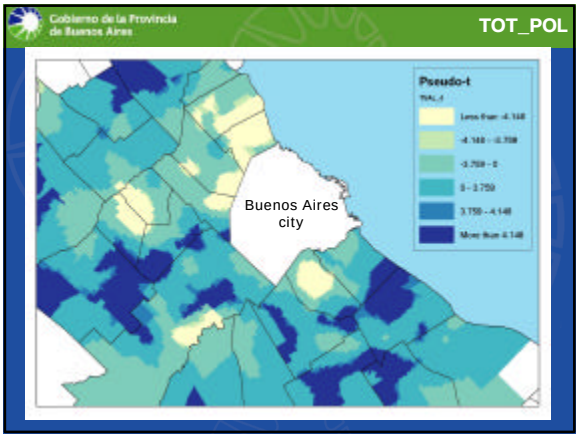
GWR Bonferroni Intervals

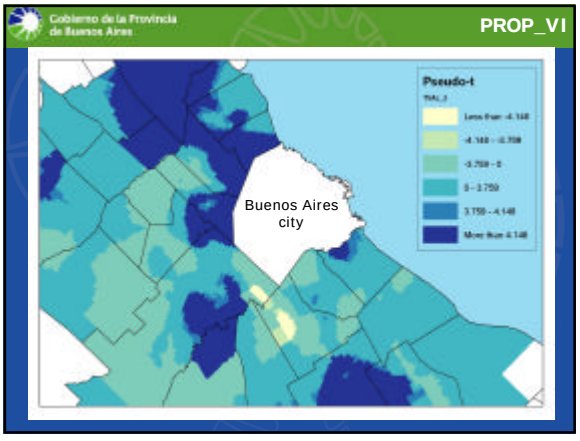
3.759 and 4.148 are the Bonferroni adjusted critical values of the t distribution at the 95 % and 99 % significance levels, respectively.

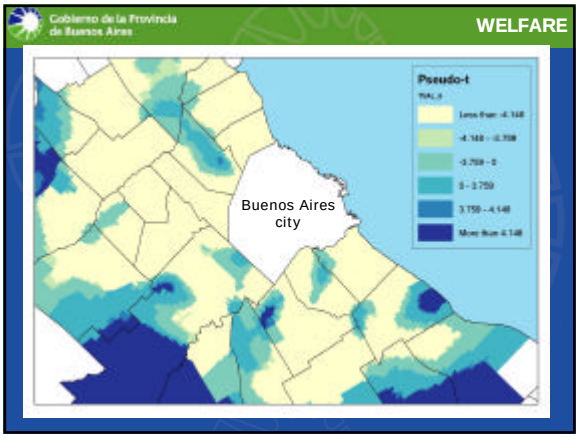

Gobierno de la Provincia de Buenos Aires
Intercept
Pseudo-t
Buenos Aires city


Gobierno de la Provincia de Buenos Aires
TOT_P
Pseudo-t
Buenos Aires city


Gobierno de la Provincia de Buenos Aires
CONF
Pseudo-t
Buenos Aires city

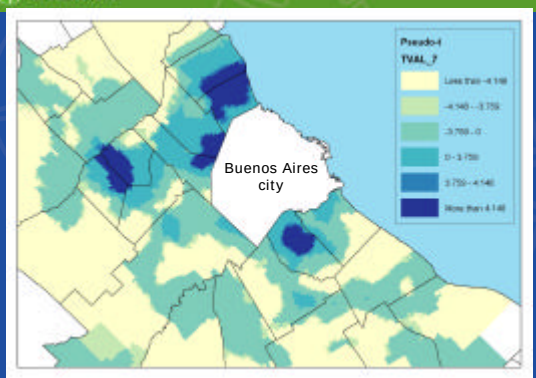

Gobierno de la Provincia de Buenos Aires
TOT_POL
Pseudo-t
Buenos Aires city


Gobierno de la Provincia de Buenos Aires
PROP_VI
Pseudo-t
Buenos Aires city


Gobierno de la Provincia de Buenos Aires
WELFARE
Pseudo-t
Buenos Aires city

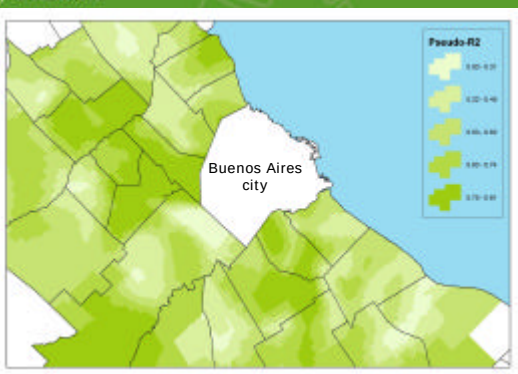
Gobierno de la Provincia de Buenos Aires

SUP_RAD



Gobierno de la Provincia de Buenos Aires

Pseudo R^2



Gobierno de la Provincia de Buenos Aires

Conclusions

- GWR method provides an excellent way to deal with spatial data.
  - The ideas behind this technique is almost straight forward to "explain" to the Law Enforcement Audience.
  - Due to our experience, we are extremely satisfied with the result.
  - R software & the GWR allows us to implement it almost effortlessly.

Gobierno de la Provincia de Buenos Aires

Conclusions

- In this example, the GWR method allows us to study local relationships between the offender's home and several predictors, and provided enough information to indicate that settlement areas can not be targeted as the sole cause for violent offender's locations. Moreover it show us that some areas with low welfare indicators and high presence of settlements expel violent offenders…

Gobierno de la Provincia de Buenos Aires

Thanks for your time

Gobierno de la Provincia de Buenos Aires